CHAPTER 4

Issues for risk adjustment in Medicare Advantage

C H A P T E R 4

Issues for risk adjustment in Medicare Advantage

Chapter summary

Health plans that participate in the Medicare Advantage (MA) program receive monthly capitated payments for each Medicare enrollee. Each capitated payment is the product of two general parts: a base rate, which reflects the payment if an MA enrollee has the health status of the national average beneficiary, and a risk score, which indicates how costly the enrollee is expected to be relative to the national average beneficiary. The purpose of the risk scores is to adjust MA payments so that they accurately reflect how much each MA enrollee would be expected to cost. In this chapter, we examine the performance of the risk-adjustment system in the MA program and offer alternatives for improving it.

Improving payment accuracy of the CMS–hierarchical condition category model

Currently, CMS uses the CMS–hierarchical condition category (CMS–HCC) model to risk adjust each MA payment. This model uses enrollees' demographics and medical conditions collected into 70 HCCs to predict their costliness. It has been shown to be a much better predictor of a beneficiary's costliness than the demographic-based model that preceded it. Analysis of the CMS–HCC model and the demographic model indicates that the CMS–HCC model explains about 11 percent of the variation in costliness among individual beneficiaries, whereas the demographic model explains only about 1 percent (Pope et al. 2004).

In this chapter

The demographic model did not include factors that are important for predicting beneficiaries' costliness—such as conditions. Consequently, it systematically overpredicted costs for healthy beneficiaries and systematically underpredicted costs for beneficiaries in poor health. Because the CMS–HCC model includes beneficiaries' conditions as well as their demographic information, it explains more of the variation in beneficiaries' costliness and predicts costs more accurately than the demographic model. However, systematic overpredictions and underpredictions may remain under the CMS–HCC model. For example, for all beneficiaries who have the same condition, the CMS–HCC model adjusts MA payments by the same proportion. But the severity of a condition varies across beneficiaries, and those with greater severity tend to be more costly. In addition, research suggests that a minimum of 20 percent to 25 percent of the variation in beneficiaries' costliness may be predictable, so the CMS–HCC may leave half or more of the predictable variation unexplained (Newhouse et al. 1997). Therefore, for a given condition it is possible that plans can be financially advantaged or disadvantaged based on the risk profile (overall health status) of their enrollees.

To the extent that systematic prediction errors occur under the CMS–HCC model, we explored several policy options for reducing these errors:

- Add measures of income and indicators for race to the model. If beneficiaries in certain income categories or racial groups tend to have greater severity for given conditions, these additional variables could reduce prediction errors.
- Include measures for the number of conditions in the model. The cost of treating a certain condition may increase as more comorbidities are present.
- Use two years of beneficiaries' diagnoses to determine their condition categories. CMS currently uses one year of beneficiaries' diagnoses to determine their conditions, but we have found that providers often do not consistently code conditions on claims from year to year. Using two years of diagnosis data (when available) would help to fully identify beneficiaries' conditions.

Our analysis indicates that including beneficiaries' race and measures of income would not improve predictive accuracy. However, including the number of conditions would improve predictive accuracy for beneficiaries who have many conditions. Using two years of diagnoses to identify beneficiaries' conditions also would improve predictive accuracy for beneficiaries who have many conditions but to a lesser extent than adding the number of conditions. In addition, using two years of diagnoses would reduce year-to-year fluctuations in beneficiaries' risk scores, which would result in more stable revenue streams for MA plans.

Because adding the number of conditions and using two years of diagnosis data both have beneficial effects, but in different ways, we also examined the effects of adding both features to the CMS–HCC model. It resulted in more accurate predictions for high-risk beneficiaries and smaller year-to-year fluctuations in beneficiaries' risk scores.

Other issues for MA risk adjustment

On several occasions, the Commission has taken a position that payments for MA enrollees should equal what they would be expected to cost in FFS Medicare (financial neutrality) (Medicare Payment Advisory Commission 2001, Medicare Payment Advisory Commission 2002, Medicare Payment Advisory Commission 2004, Medicare Payment Advisory Commission 2005). An underlying rationale for this policy is that it encourages beneficiaries to enroll in whichever sector (MA or FFS) is more efficient in their local health care market. Two recently published papers have implications for the interaction between risk adjustment and financial neutrality. One paper shows that in FFS Medicare, regions that have high per capita service use also have high average risk scores, and areas that have low per capita service use have low average risk scores (Song et al. 2010). At least some of the regional difference in risk scores is due to differences in service use that do not reflect differences in health status; that is, risk scores are high in some regions simply because beneficiaries get more health care, not because they are sicker. If these same regional differences in service use and risk scores occur in the MA program, they drive MA payments higher in high-use regions. Adjustments could be made to eliminate these differences.

A second paper shows there are differences between FFS Medicare and a large MA plan in the relative costliness of treating conditions (Newhouse et al. 2011). If these cost differences between FFS Medicare and MA plans are widespread, risk adjustment underpredicts MA costs for some conditions and overpredicts MA costs for other conditions. An issue to consider is whether it is more appropriate for CMS to estimate the CMS–HCC model by using cost and diagnosis data from MA enrollees rather than FFS beneficiaries.

Both papers have implications for equity in the MA program: Adjusting MA risk scores to reduce the effects of regional differences in risk scores would reduce regional variations in MA payments, and using data from MA enrollees to estimate the CMS–HCC model would reduce incentives for plans to attract beneficiaries who have some conditions and avoid beneficiaries who have other conditions. However, both issues are inconsistent with the concept of financial neutrality between MA and FFS Medicare. These issues will have to be discussed in the future.

A final issue regarding risk adjustment is that analyses by CMS and the Government Accountability Office (GAO) both indicate that risk scores have increased at a faster rate in the MA program than in FFS Medicare (Centers for Medicare & Medicaid Services 2009, Government Accountability Office 2012). The higher growth rate in MA has resulted in MA enrollees having higher risk scores than they would have in FFS Medicare. In an effort to bring MA risk scores in line with those in FFS Medicare, CMS has made adjustments to MA risk scores, but GAO believes that larger adjustments should be made. ■

Health plans that participate in the Medicare Advantage (MA) program receive monthly capitated payments for each Medicare enrollee. Each capitated payment is the product of two general parts: a base rate, which reflects the payment if an MA enrollee has the health status of the national average Medicare beneficiary, and a risk score, which indicates how costly the enrollee is expected to be relative to the national average beneficiary.

Over the years, CMS has employed various methods for determining MA enrollees' risk scores. Currently, CMS uses the CMS–hierarchical condition category (CMS–HCC) risk-adjustment model, which uses enrollees' demographics and condition categories (such as diabetes and stroke) to predict their costliness. The demographic variables include age, sex, Medicaid status, institutional status, eligibility based on being disabled, and eligibility based on age but originally eligible because of disability.

All demographic variables are from the year for which beneficiaries' costs are to be predicted (the prediction year). The condition categories are based on diagnoses recorded on physician, hospital outpatient, and hospital inpatient claims in the year before beneficiaries' costs are to be predicted (the base year). This makes the CMS–HCC a prospective model, as opposed to a concurrent model, which would use conditions from the prediction year. It is logical to use a prospective model in the MA program because the express purpose of MA plans is to provide care to manage their enrollees' conditions. If concurrent risk adjustment were used, MA plans would be reimbursed as their enrollees' conditions occur, rather than being paid to manage existing chronic conditions.

An underlying feature of a prospective risk-adjustment model for beneficiaries with a given set of conditions is that it underpredicts costs for some beneficiaries, overpredicts for others, but predicts accurately on average. However, when prediction inaccuracies occur systematically with identifiable beneficiary characteristics, plans can benefit if their enrollees have characteristics predictive of lower-than-average costs (favorable selection) or be disadvantaged if their enrollees have characteristics predictive of higher-than-average costs (adverse selection). An ideal risk-adjustment system would eliminate all opportunities for favorable selection, but a risk adjuster can be less than ideal and still be effective if it makes efforts by plans to identify favorable risks prohibitively costly.

To the extent that favorable selection occurs in the MA program, it could be caused by the behavior of plans or beneficiaries. Plans can attract favorable risks by structuring benefits that are attractive to relatively healthy beneficiaries or by marketing their products so that they attract healthy enrollees. Alternatively, relatively healthy beneficiaries may find the structure of managed care plans more attractive than do beneficiaries in poor health.

Selection problems can be reduced by improving risk adjustment to reduce the extent of systematic prediction errors. An alternative method for reducing selection problems is partial capitation, which would pay plans partly on the basis of capitated rates and partly on the actual costs of providing care. This would reduce the likelihood of plans experiencing large losses from very sick enrollees or large profits from healthy enrollees. However, it is not clear what fraction of the payments should be capitated and what fraction should be based on costs.

For each MA enrollee, CMS obtains from the enrollee's plan the condition codes from encounters with physicians, hospital outpatient departments, and hospital inpatient departments. CMS maps the condition codes into hierarchical condition categories (HCCs), which define broad condition categories, such as diabetes and congestive heart failure. All condition codes fall into one of the 189 CMS-defined HCCs. However, CMS uses only 70 HCCs in the CMS–HCC model, so many conditions have no effect on beneficiaries' risk scores.[1] Some conditions, such as diabetes and cancer, are actually represented by groups of several HCCs, which differ according to severity. CMS has determined that a beneficiary cannot have more than one HCC indicated in each of these condition groups. If CMS finds that a beneficiary has conditions that map into more than one HCC within a condition group, only the highest cost HCC is used in predicting the beneficiary's costliness. For example, the CMS–HCC model has five diabetes HCCs. If CMS finds that a beneficiary has condition codes that fit the HCC "diabetes with acute complications" and the HCC "diabetes without complications," CMS drops the HCC "diabetes without complications."

CMS calibrates the additional costliness associated with each demographic variable and each HCC in the model using cost, demographic, and diagnosis data from beneficiaries in fee-for-service (FFS) Medicare, but CMS has begun collecting data on MA beneficiaries and intends to use those data to calibrate the CMS–HCC model.[2] CMS applies linear regression methods to obtain coefficients on each variable in the model. If a beneficiary has a particular variable represented in the CMS–HCC model, the coefficient on that variable indicates its marginal cost.

A model that CMS used before the CMS–HCC model included only beneficiaries' demographic data (demographic model). The demographic model does not include important observable characteristics of beneficiaries that affect their costliness, such as medical conditions. Consequently, the demographic model explains a small fraction of the variation in beneficiaries' Medicare costliness (about 1 percent) and, within a given demographic category, systematically overpredicts costs for relatively healthy beneficiaries and underpredicts costs for the sickest beneficiaries, leaving the potential for selection problems.

The CMS–HCC model has been shown to be a much better predictor of a beneficiary's costliness. For example, it explains about 11 percent of the variation in beneficiaries' costliness. Therefore, the CMS–HCC model likely mitigates selection problems in the MA program.

However, the CMS–HCC model has shortcomings such that it may not have fully eliminated systematic prediction inaccuracies:

- Research on variation in individual-level health care spending suggests that at least 20 percent to 25 percent of the variation in spending can be predicted, with the remaining being random and, hence, unpredictable (Newhouse et al. 1997). Because the CMS–HCC model explains about 11 percent of the variation in spending, it may leave half or more of the predictable variation unexplained.
- For all enrollees with a given health condition, the CMS–HCC model adjusts MA capitated payments by the same rate. For example, the CMS–HCC model increases capitated payments for all MA enrollees with acute myocardial infarction by 35.9 percent above the base rate. However, within condition categories, some beneficiaries are healthier and less costly than others, while some are sicker and more costly.

Because of these shortcomings of the CMS–HCC model, there is a potential for MA plans to benefit financially if they have a relatively healthy beneficiary profile or to be disadvantaged if they have a sicker beneficiary profile. This is especially relevant to plans that specialize in managing the care for the sickest beneficiaries, such as special needs plans (SNPs) and plans in the Program of All-Inclusive Care for the Elderly (PACE), because payments may not be adequately adjusted to effectively provide care.

We have done an analysis that suggests that MA enrollees are healthier and less costly than their FFS counterparts, meaning that favorable selection may be occurring in the MA program. In this chapter, we discuss this analysis as well as options for modifying the CMS–HCC model to mitigate systematic prediction errors.

This chapter discusses two other issues concerning risk adjustment in the MA program:

- Research indicates that geographic differences in per capita service use in FFS Medicare lead to geographic differences in risk scores that do not reflect differences in health status among FFS beneficiaries (Song et al. 2010). If regional differences in service use also lead to regional differences in risk scores in the MA program, then plans in regions where service use is high have higher risk scores, which drive capitated payments in those regions above the level in lower use regions.
- The coefficients on the conditions in the CMS–HCC model indicate the relative costliness of treating those conditions in FFS Medicare. However, the CMS–HCC model is used to risk-adjust payments in the MA program, where the relative costs of treating conditions may differ from costs in FFS Medicare (Newhouse et al. 2011). This raises questions of whether it is more appropriate to continue to calibrate the CMS–HCC model using data on FFS beneficiaries, or to switch to data on MA enrollees.

Evidence that MA enrollees tend to be lower cost than FFS beneficiaries

Recently, there has been renewed interest in examining the extent to which favorable selection occurs in the MA program. One study found a substantial amount of favorable selection (Brown et al. 2011). Another study used cost data from a large MA plan and found little correlation between how costly a condition is to treat in FFS Medicare and the extent to which beneficiaries with that condition are profitable to that plan (Newhouse et al. 2011). This gives the plan little incentive to try to select against the sickest, highest cost beneficiaries.

Within a given HCC, the severity of the condition and hence the cost of treating it can vary. For example, we examined FFS beneficiaries who were grouped into the HCC for congestive heart failure (CHF) in 2008 and had no other HCCs. In 2008, the beneficiary at the 95th percentile of costliness had more than $37,000 in Medicare spending, while the beneficiary at the 5th percentile had $115 in Medicare spending. Despite these large cost differences for beneficiaries who have the same condition, the CMS–HCC model adjusts the payment rate for each beneficiary who has CHF by the same proportion (41 percent). Therefore, it is possible that, for beneficiaries in a given HCC, some will be profitable to MA plans because they are low-severity cases while others will not be profitable because they are high-severity cases.

If beneficiaries who have the same condition are randomly selected into MA plans in sufficiently large numbers, those who are profitable will be offset by those who are unprofitable, resulting in no financial gain or loss for the plan. However, if the selection of these beneficiaries is not random, it is possible that those who enroll in MA plans are on average profitable. This could occur either through the actions of plans—perhaps through benefits that are attractive to healthier beneficiaries or marketing techniques that target those beneficiaries—or because relatively healthy beneficiaries find the structure of MA plans more attractive than do sicker beneficiaries.

We conducted a study using two measures that may suggest, but not confirm, whether MA enrollees are, on average, lower risk than FFS beneficiaries. We divided the beneficiaries who were in FFS Medicare in 2007 into two groups: those who stayed in FFS Medicare in 2008 (stayers) and those who enrolled in MA plans in 2008 (joiners). For each group, we calculated the mean FFS costliness in 2007 in each HCC and used beneficiaries' risk scores to adjust for differences in health status. We reasoned that, for each HCC, if the risk-adjusted mean cost of the joiners was below that of the stayers, it indicates that lower cost beneficiaries are more likely to enroll in the MA program. We also identified the beneficiaries who were in FFS Medicare in 2008 and divided them into two groups: those who were in FFS Medicare throughout 2007 and those who left an MA plan in 2007. For both groups, we calculated the mean FFS costliness in 2008 of the beneficiaries who had conditions in each HCC and used beneficiaries' risk scores to adjust for differences in health status. We reasoned that, for each HCC, if the risk-adjusted mean cost of the beneficiaries who left an MA plan in 2007 was above that of beneficiaries who were in FFS Medicare throughout 2007 and 2008, it indicates that beneficiaries who stay in MA plans tend to be lower cost than those who leave MA for FFS Medicare.

Our results from both analyses suggest that MA enrollees are, on average, lower cost than FFS beneficiaries. In the first analysis, for 68 of the 70 HCCs, beneficiaries who joined an MA plan in 2008 had lower FFS costs in 2007 than the beneficiaries who stayed in FFS Medicare in 2008. On average, the joiners had costs that were 15 percent lower than the stayers. In the second analysis, beneficiaries who left an MA plan in 2007 had FFS costs in 2008 that averaged 16 percent higher than beneficiaries who were in FFS Medicare throughout 2007. Moreover, for 69 of the 70 HCCs, beneficiaries who disenrolled from MA in 2007 had higher average costs in 2008 than those who were in FFS Medicare throughout 2007.

Although these results suggest that MA enrollees are lower cost than FFS beneficiaries, we emphasize that they are not conclusive. It is possible that beneficiaries with relatively low costs are more likely to enroll in an MA plan but that their costs increase after enrollment. Possible reasons this may occur include that their costs regress to the mean over time, they lacked supplemental coverage while in FFS Medicare, or they have low incomes and the more comprehensive coverage that often occurs in MA plans encourages them to increase their service use.

Improving predictive accuracy of the CMS–HCC model

We have evaluated three alternatives for improving the predictive accuracy of the CMS–HCC model so that systematic prediction errors are reduced. All three options involve using more data than are currently used in the CMS–HCC model:

- Add socioeconomic variables such as race/ethnicity and income to the model. This model includes all variables in the current CMS–HCC model plus race/ethnicity indicators (African American, Hispanic, White, other race) and income level, which we approximated by the per capita income in the beneficiary's county of residence.
- Add indicators for the number of conditions beneficiaries have. This model includes all variables in the current CMS–HCC model plus indicators of whether beneficiaries have zero, one, two, three, four, or five or more HCCs.[3]
- Use two years of diagnosis data (when available) to determine each beneficiary's HCCs rather than one

TABLE 4–1

Adding indicators of race and measure of income has little effect on predictive accuracy of CMS–HCC model

Category	Predictive ratio: Standard model	Predictive ratio: Race/income model
Specific conditions		
Diabetes	1.00	1.00
COPD	1.01	1.01
CHF	0.99	0.99
Cancer	0.99	0.99
Mental illness	1.00	1.00
Schizophrenia	1.00	1.00
AMI	1.03	1.02
Unspecified stroke	1.01	1.00
All strokes	1.01	1.00
Number of conditions		
0	0.94	0.94
1	1.02	1.02
2	1.03	1.03
3	1.03	1.02
4	1.02	1.02
5 or more	0.98	0.98
8 or more	0.95	0.94

Note: CMS–HCC (CMS–hierarchical condition category), COPD (chronic obstructive pulmonary disease), CHF (congestive heart failure), AMI (acute myocardial infarction). We determined the number of conditions by counting the number of HCCs a beneficiary maps into. Both models use one year of diagnosis data to determine beneficiaries' conditions.

Source: MedPAC analysis of 5 percent standard analytic claims files and 5 percent Medicare denominator file.

year of diagnosis data, which CMS currently uses. Obviously, two years of diagnosis data would not be available for beneficiaries in their first or second year of Medicare eligibility. For those in the first year of Medicare eligibility, the demographic model that CMS currently uses for new enrollees is a viable option. For those in the second year of eligibility, we could use the current version of the CMS–HCC model.

Adding socioeconomic variables does not improve predictive accuracy of CMS–HCC model

We calibrated a model that has all the variables of the current version of the CMS–HCC model, which has 70 HCCs (standard model). We also calibrated a version that has the same variables as the standard model plus race/ethnicity indicators (African American, Hispanic, White, and other race) and income level, which we approximated by the per capita income in the beneficiary's county of residence (race/income model). For a description of the method we used to calibrate the models presented in this chapter, see the text box (pp. 106–107).

We evaluated the predictive accuracy of the standard and race/income versions using two measures:

- R^2, a statistical measure of how much of the variation in costliness among individuals is explained by the model: The closer the R^2 is to 1.0, the more of the variation the model has explained.
- Predictive ratio: For a group of beneficiaries, it is the total costliness predicted by the model divided by the total actual costliness of the group. The closer the predictive ratio is to 1.0, the better the model has predicted the actual costs. Predictive ratios less than 1.0 indicate the predicted costs are below the actual costs (underprediction); predictive ratios greater than 1.0 indicate the predicted costs are above actual costs (overprediction).

The R^2 gives a general sense of how well a model accounts for variations in costs across individuals. However, strategies to attract favorable risks are typically based on characteristics such as conditions that define groups of beneficiaries, not on specific individuals. Therefore, many analysts prefer to use predictive ratios to evaluate the predictive accuracy of risk-adjustment models (Frogner et al. 2011, Pope et al. 2004, Pope et al. 2011). For beneficiaries with a given attribute, the predictive ratio indicates (on average) the extent to which a model overpredicts or underpredicts the costliness of the beneficiaries with the attribute and by how much.

The addition of race and income variables to the standard model did very little to enhance its predictive accuracy. Using the standard CMS–HCC model, we obtained an R^2 of 0.1100. Adding race and income variables had no effect on the R^2.

We also used predictive ratios to examine how accurately these two models predict beneficiaries' costliness for nine condition categories. For most of these conditions, both models predict beneficiaries' costliness quite well, but they overpredict costs to some degree for acute myocardial infarction (AMI). Within each of these conditions—as well as all conditions represented in the CMS–HCC model—some beneficiaries are relatively healthy and have no other conditions, while other beneficiaries are much sicker and have many other conditions. We analyzed categories of beneficiaries identified by number of conditions: zero, one, two, three, four, five or more, and eight or more. We found that both models underpredict costliness to some degree for beneficiaries who have five or more conditions and by a larger degree for those who have no conditions or eight or more conditions. Also, both models overpredict costliness to some degree for categories defined by one condition, two conditions, three conditions, and four conditions (Table 4-1).

These prediction errors can be seen when we parse the beneficiaries who have diabetes by how many other conditions they have. For diabetics who have one other condition (and would be in the two conditions group in Table 4-1), the predictive ratio is 1.03; for diabetics who have at least seven other conditions (and would be in the eight or more conditions group in Table 4-1), the predictive ratio is 0.93. However, these deviations from 1.0 are a concern only if there is systematic selection into MA plans of the beneficiaries who are in categories for which the predictive ratio is above 1.0.

The important points to take away from Table 4-1 are:

- The CMS–HCC model accurately predicts costs, on average, for most conditions that are represented in the model.
- However, among beneficiaries who have the same condition, some are relatively healthy and have no other conditions or only a few other conditions, while others are sicker and have many other conditions. For those who have only a few conditions, the CMS–HCC model slightly overpredicts costs, and for those who have many conditions, the model underpredicts costs, and the underprediction increases as the number of conditions increases. Consequently, SNPs and plans in PACE, which focus on the sickest beneficiaries, may be at a disadvantage, while plans that are able to attract the healthiest beneficiaries with a given condition may benefit.
- Adding race and income to the CMS–HCC model does little to improve the model's predictive accuracy.

Including number of conditions improves predictive accuracy

We used 2007 diagnosis data and 2008 demographic and program cost data to calibrate two versions of the CMS–HCC model. One is the standard model that CMS uses in the MA program and is the same standard model in Table 4-1. The other is the conditions model, which adds to the standard model six indicators for how many conditions each beneficiary has, as determined by the beneficiary's diagnoses: zero, one, two, three, four, and five or more conditions. We define number of conditions as the number of HCCs that each beneficiary's conditions map into.

This standard model has an R^2 of 0.1100, indicating that it explains 11 percent of the variation in beneficiaries' Medicare costs. When we add the six measures indicating the number of conditions, the improvement in the R^2 is nearly imperceptible, 0.1105.

We also calculated predictive ratios for nine condition categories using both the standard and conditions models. For all nine conditions, the predictive ratios show little or no change between the two models. For most conditions, both predict reasonably well, with the exception being AMI, where there is some degree of overprediction under both models. As we saw in Table 4-1, the standard model underpredicts for beneficiaries who have zero conditions, five or more conditions, and eight or more conditions and overpredicts for one, two, three, and four conditions. In contrast, the conditions model predicts quite accurately for each of those groups (Table 4-2, p. 104). Because the conditions model predicts accurately for the sickest beneficiaries (those who have many conditions), it may be beneficial for SNPs and PACE plans.

Using two years of diagnosis data stabilizes risk scores and improves predictive accuracy

Previous research indicates that in FFS Medicare, a beneficiary who has a chronic condition indicated on a claim in one year often will not have that condition appear on a claim in the following year (Frogner et al. 2011, Medicare Payment Advisory Commission 1998). If this inconsistent coding of beneficiaries' chronic conditions also occurs among MA plans, beneficiaries' risk scores will often have large year-to-year changes.

We evaluated the extent to which beneficiaries who were coded for the HCCs for kidney failure, stroke, quadriplegia or paraplegia, diabetes, CHF, and chronic obstructive pulmonary disease in 2007 also were coded for those HCCs in 2008. We did this for both FFS enrollees and MA enrollees.

Our results indicate that coding for all of these conditions was not consistent from year to year. The same was true for beneficiaries in both FFS Medicare and MA plans

TABLE 4–2

Adding number of conditions to CMS–HCC model improves predictive accuracy for beneficiaries who have many conditions

Category	Predictive ratio: Standard model	Predictive ratio: Conditions model
Specific conditions		
Diabetes	1.00	1.00
COPD	1.01	1.01
CHF	0.99	0.99
Cancer	0.99	0.99
Mental illness	1.00	1.00
Schizophrenia	1.00	1.00
AMI	1.03	1.03
Unspecified stroke	1.01	1.01
All strokes	1.01	1.00
Number of conditions		
0	0.94	1.00
1	1.02	1.00
2	1.03	1.00
3	1.03	1.00
4	1.02	1.00
5 or more	0.98	0.99
8 or more	0.95	1.00

Note: CMS–HCC (CMS–hierarchical condition category), COPD (chronic obstructive pulmonary disease), CHF (congestive heart failure), AMI (acute myocardial infarction). We determined the number of conditions by counting the number of HCCs a beneficiary maps into. Both models use one year of diagnosis data to determine beneficiaries' conditions.

Source: MedPAC analysis of 5 percent standard analytic claims files and 5 percent Medicare denominator file.

(Table 4-3). This lack of consistent coding over time presents two problems for risk adjustment. First, in a given year, many FFS beneficiaries who have a condition will not have that condition appear on a claim. Because CMS uses conditions recorded on claims for FFS beneficiaries to calibrate the CMS–HCC model, the model may not accurately reflect the true additional cost of a particular condition. Second, inconsistent coding of conditions in MA results in greater year-to-year fluctuations in enrollees' risk scores, which leads to less stable payments and revenue streams to MA plans.

These problems related to inconsistent coding of conditions would be mitigated if CMS used two years of beneficiaries' diagnosis data rather than one year to calibrate the CMS–HCC model and determine beneficiaries' risk scores. The Commission has recommended this position in the past (Medicare Payment Advisory Commission 2000).

We calibrated a version of the CMS–HCC model that is the same as the standard model, but we used two years of diagnosis data to assign beneficiaries to HCCs (two-year model). We found that this model produces risk scores that are more consistent over time than does the standard CMS–HCC model. For example, we found that the correlation coefficient between the 2008 and 2009 risk scores for more than 1 million beneficiaries was 0.62 using the standard model and 0.80 using the two-year model, where the correlation coefficient indicates how strongly one variable is correlated with another. The closer a correlation coefficient is to 1.0, the more closely two variables are correlated.

We also found that for specific conditions, there is little difference in predictive accuracy between the standard model and the two-year model, except for mental illness. However, the two-year model predicts more accurately for beneficiaries who have five or more conditions and for those who have eight or more conditions (Table 4-4). As we mentioned earlier, this means that for most conditions, both models pay accurately, on average, except for AMI, where there is some degree of overprediction of costs. However, for those who have five or more conditions and those who have eight or more conditions, the two-year model underpredicts by a lesser amount than does the

TABLE 4–3

Beneficiaries who had chronic condition on claim in 2007 often did not have same condition on claim in 2008

Condition category	Of those with condition coded in 2007, percent who did not have it coded again in 2008: FFS Medicare	Of those with condition coded in 2007, percent who did not have it coded again in 2008: MA program
Diabetes	12.9%	10.9%
COPD	33.8	29.9
CHF	37.9	34.4
Kidney failure	35.4	28.9
Stroke	56.7	59.0
Quadriplegia/paraplegia	58.7	62.3

Note: FFS (fee-for-service), MA (Medicare Advantage), COPD (chronic obstructive pulmonary disease), CHF (congestive heart failure).

Source: MedPAC analysis of 2007 and 2008 risk score files from Acumen, LLC, and 2006 and 2007 Medicare denominator files from Acumen, LLC.

TABLE 4–4

Using two years of diagnoses in CMS–HCC model improves predictive accuracy for beneficiaries who have many conditions

Category	Predictive ratio: Standard model	Predictive ratio: Two-year model
Specific conditions		
Diabetes	1.00	1.00
COPD	1.01	1.01
CHF	0.99	1.00
Cancer	0.99	0.99
Mental illness	1.00	0.96
Schizophrenia	1.00	1.00
AMI	1.03	1.02
Unspecified stroke	1.01	1.01
All strokes	1.01	1.01
Number of conditions		
0	0.94	0.92
1	1.02	1.00
2	1.03	1.02
3	1.03	1.03
4	1.02	1.03
5 or more	0.98	1.00
8 or more	0.95	0.97

Note: CMS–HCC (CMS–hierarchical condition category), COPD (chronic obstructive pulmonary disease), CHF (congestive heart failure), AMI (acute myocardial infarction). We determined the number of conditions by counting the number of HCCs a beneficiary maps into. The standard model uses one year of diagnosis data to determine beneficiaries' conditions, the two-year model uses two years of diagnosis data.

Source: MedPAC analysis of 5 percent standard analytic claims files and 5 percent Medicare denominator file.

standard model. In summary, the two-year model offers the advantages of smaller year-to-year fluctuations in risk scores and more accurate payments for the sickest beneficiaries.

Including number of conditions and using two years of diagnosis data have the benefits of both

We also analyzed the effects of a version of the CMS–HCC model that includes indicators for number of conditions and uses two years of diagnosis data to determine beneficiaries' HCCs (combined model). The combined model has the benefits of both the conditions model and the two-year model: It improves the predictive accuracy for beneficiaries who have many conditions and it reduces year-to-year fluctuations in beneficiaries' risk scores (Table 4-5). However, the combined model underpredicts costliness for mental illness, which also occurs under the two-year model.

Issues related to financial neutrality between FFS Medicare and the MA program

CMS estimates the CMS–HCC model using cost, demographic, and diagnosis data from FFS beneficiaries. Therefore, the coefficients for each HCC indicate the relative costliness of treating those conditions in FFS Medicare. On several occasions, the Commission has

TABLE 4–5

Adding measures for number of conditions and using two years of diagnoses in CMS–HCC model improves predictive accuracy for beneficiaries who have many conditions

Category	Predictive ratio: Standard model	Predictive ratio: Combined model
Specific conditions		
Diabetes	1.00	1.00
COPD	1.01	1.01
CHF	0.99	1.00
Cancer	0.99	0.99
Mental illness	1.00	0.95
Schizophrenia	1.00	1.00
AMI	1.03	1.02
Unspecified stroke	1.01	1.01
All strokes	1.01	1.01
Number of conditions		
0	0.94	1.01
1	1.02	1.00
2	1.03	1.00
3	1.03	1.01
4	1.02	0.99
5 or more	0.98	0.99
8 or more	0.95	1.00

Note: CMS–HCC (CMS–hierarchical condition category), COPD (chronic obstructive pulmonary disease), CHF (congestive heart failure), AMI (acute myocardial infarction). We determined the number of conditions by counting the number of HCCs a beneficiary maps into. The standard model uses one year of diagnosis data to determine beneficiaries' conditions, the combined model uses two years of diagnosis data.

Source: MedPAC analysis of 5 percent standard analytic claims files and 5 percent Medicare denominator file

Methods used in regression analysis

For this chapter, we estimated several versions of the CMS–hierarchical condition category (CMS–HCC) risk-adjustment model. We used the same general method to produce all of the regression-based results presented. The only differences between regressions are the explanatory variables. The results in all our regressions are based on a 5 percent sample of fee-for-service (FFS) beneficiaries.

In each regression, we used data from 2007 and 2008. The 2007 data are the HCCs based on diagnoses from hospital inpatient, hospital outpatient, and physician claims that we used to determine each beneficiary's condition categories for 2008, which are defined by 70 HCCs in the CMS–HCC model. Examples of conditions defined by the HCCs include diabetes with various degrees of severity, congestive heart failure (CHF), and chronic obstructive pulmonary disease (COPD). The 2008 data include the following for each beneficiary: total costliness to the Medicare program, age, sex, Medicaid status, whether institutionalized for three consecutive months, and whether eligible for Medicare on the basis of age but originally eligible because of disability.

To be included in the analysis, beneficiaries had to meet the following requirements: in both Part A and Part B of Medicare throughout 2007, in both Part A and Part B throughout their Medicare eligibility in 2008, no Medicare Advantage enrollment at any time in 2007 or 2008, no hospice care in 2008, not classified as having end-stage renal disease in 2008, lived within the United States for all of 2007 and 2008, no Medicare as a secondary payer in 2007 or 2008, and not long-term institutionalized in 2008. In addition, the results in Tables 4-4 and 4-5 include analyses of versions of the CMS–HCC model that use two years of diagnosis data to determine each beneficiary's HCCs. For this regression, we used diagnoses from 2006 and 2007 claims; when we used data from 2007 to exclude beneficiaries in the other regressions, we used data from 2006 and 2007 to exclude beneficiaries in the two-year regression.

In each regression, the dependent variable was each beneficiary's 2008 costliness to FFS Medicare that we annualized if the beneficiary was in FFS Medicare for only a fraction of 2008. That is, we divided each beneficiary's 2008 costliness to FFS Medicare by the fraction of the year the beneficiary was in FFS Medicare in 2008. Each regression included the following explanatory variables:

- 70 HCCs;
- 24 categories indicating age and sex;
- 4 variables indicating Medicaid status;

(continued next page)

stated that payments to MA plans should be equal to what MA enrollees would cost in FFS Medicare (financial neutrality) (Medicare Payment Advisory Commission 2001, Medicare Payment Advisory Commission 2002, Medicare Payment Advisory Commission 2004, Medicare Payment Advisory Commission 2005). The current method of using cost and diagnosis data from FFS Medicare to estimate the CMS–HCC model is consistent with the goal of financial neutrality.[4]

In light of two recently published papers, more discussion about financial neutrality between FFS Medicare and the MA program may be appropriate. One study found that large regional differences in service use among FFS beneficiaries lead to large regional differences in risk scores (Song et al. 2010). This study indicates that these regional differences in risk scores are due, at least in part, simply to differences in service use rather than differences in health status. It is not known if these regional differences in risk scores also occur in the MA program. If they do, higher risk scores in high-use regions will drive up MA payments not because MA enrollees are less healthy but simply because service use is higher.

A second study found that the relative cost of treating specific conditions differs widely between a large MA plan and FFS Medicare. For some conditions, the relative cost is

Methods used in regression analysis (continued)

- 2 variables indicating beneficiaries who are eligible because of age but were originally eligible because of disability;
- 5 categories indicating that beneficiaries are disabled and have 1 of 5 conditions: opportunistic infections, severe hematologic disorders, drug or alcohol psychosis, drug or alcohol dependence, and cystic fibrosis; and
- 6 disease interaction terms: diabetes and CHF, diabetes and cardiovascular disease, CHF and COPD, CHF with COPD and coronary artery disease, renal failure (RF) and CHF, and RF with CHF and diabetes.

These are the same dependent and explanatory variables that CMS includes in the version of the CMS–HCC model it currently uses.

We ran six weighted regressions to produce the results in Tables 4-1, 4-2, 4-4, and 4-5, where the weight is the fraction of the year each beneficiary was in FFS Medicare in 2008, which is the same fraction we used to annualize each beneficiary's costs. For Table 4-1, we ran a standard version of the CMS–HCC model, which includes all the variables listed above, and we ran a race/income version that includes the same variables plus indicators for each beneficiary's race (African American, Hispanic, White, or other) and income, which we approximated by the per capita income in each beneficiary's county of residence.

For Tables 4-2, 4-4, and 4-5, we ran the following regressions:

- a conditions version, which includes all the variables in the standard version plus indicators of whether each beneficiary has zero, one, two, three, four, or five or more HCCs (Table 4-2);
- a two-year version, which includes all the variables in the standard version, but HCCs for each beneficiary are based on two years of diagnosis data rather than the single year used in the standard version (Table 4-4); and
- a version that combines the conditions version and the two-year version and includes all the variables in the standard version, whether each beneficiary has zero, one, two, three, four, or five or more HCCs, and HCCs that are based on two years of diagnosis data (Table 4-5).

We developed the analytic samples for each of these regressions from 5 percent samples of all FFS beneficiaries in 2008. It resulted in analytic samples of about 1.2 million beneficiaries for the standard and conditions versions and about 1.1 million beneficiaries for the two-year version and the version that combines the conditions and two-year versions. ■

lower in the MA plan; for other conditions, the relative cost in the MA plan is higher (Newhouse et al. 2011).

Should MA risk scores be adjusted for regional differences in service use?

Risk adjustment affects payments to MA plans through two mechanisms. First, county-level benchmarks depend directly on each county's per capita FFS spending, divided by the county's average CMS–HCC risk score among FFS beneficiaries. Second, CMS uses the risk scores to adjust MA payments for each enrollee.

CMS–HCC risk scores depend heavily on beneficiaries' conditions that providers have coded on claims. Research indicates that among FFS beneficiaries, per capita service use is higher in some areas of the country than in others. Moreover, average risk scores among FFS beneficiaries are highest in regions where service use is highest and lowest in regions where service use is lowest (Song et al. 2010).

This correlation between regional differences in service use and regional differences in risk scores could occur for two reasons. First, it could occur because those in the high-use regions are sicker. In this case, the relatively high risk scores in the high-use regions accurately reflect regional differences in health status. Second, it could occur because beneficiaries in high-use areas simply use more medical care without being sicker than beneficiaries in lower use areas. In this case, the relatively high risk scores in high-use areas do not reflect regional differences in health status. Research indicates that at least part of the regional differences in risk scores is due to beneficiaries in high-use regions using more medical care without actually being sicker. That is, beneficiaries in high-use areas would have lower risk scores if they lived in regions where service use was lower (Song et al. 2010).

To the extent that regional differences in service use cause differences in risk scores, there may be little effect on the county-level benchmarks. For example, if a county has a high level of service use, it is likely to be reflected in both high per capita FFS spending and a high average risk score among FFS beneficiaries. The high FFS spending and high average risk score should largely offset each other, so the county benchmark should be unaffected.

Regional differences in service use are more likely to have an effect on MA payments through the risk scores of MA enrollees. In regions with relatively high service use among MA enrollees, it is possible that providers' coding of conditions is more intensive than in other regions, leading to higher risk scores and MA payments in those regions. However, it is not clear whether the regional differences in service use and risk scores that occur in the FFS program also occur in the MA program because data are not available to make that determination. But, CMS has begun collecting beneficiary-level cost and diagnosis data from MA plans; after it has collected multiple years of these data, it may be possible to replicate the analysis by Song and colleagues for the MA population.

Addressing regional differences in risk scores due to differences in service use

The study by Song and colleagues (2010) divided the country into regions and determined per capita service use in FFS Medicare in each region. The regions were sorted into quintiles of per capita service use. The authors found that regional differences in service use led to regional differences in how intensively conditions are coded on claims, which resulted in average risk scores in the highest quintile of service use that were 15 percent higher than they would have been had the beneficiaries lived in a region in the lowest quintile.

If there are similar regional differences in the MA program, adjustments could be made to MA risk scores. Using the results from Song and colleagues as a hypothetical example, these adjustments could work as follows. Regions that are in the middle (third) quintile of service use could be used as the baseline. MA enrollees residing in a region that is in the third quintile would have no adjustment to their risk scores. Relative to the third quintile, Song and colleagues found that regional differences in service use result in risk scores that are 5.2 percent lower in the first quintile, 1.7 percent lower in the second quintile, 5.7 percent higher in the fourth quintile, and 8.8 percent higher in the fifth quintile. MA risk scores in these four regions could be adjusted by these percentages to account for differences in coding.

Adjusting risk scores for MA enrollees for the effects of regional differences in service use should be considered alongside the Commission's previously stated position on financial neutrality between the MA and FFS programs. On the one hand, making regional adjustments to MA risk scores is somewhat inconsistent with financial neutrality because one sector (FFS or MA) would have a financial advantage over another sector for treating the same patient or condition. On the other hand, to the extent that regional differences in service use that are independent of beneficiaries' health status result in regional differences in risk scores among MA enrollees, plans in high-use regions would have higher payments than plans in lower use regions simply because of regional differences in use rates. It may be appropriate to discuss the merits of each alternative.

Should CMS use FFS or MA data to estimate the CMS–HCC model?

If the large differences found by Newhouse and colleagues between the cost of treating conditions in a large MA plan and FFS Medicare also occur in most or all other MA plans, the CMS–HCC model underpredicts the costs in MA plans for some conditions and overpredicts the costs in MA plans for other conditions. Obviously, it is financially beneficial for plans to have beneficiaries who have conditions for which costs are overpredicted and avoid beneficiaries who have conditions for which costs are underpredicted.

In light of the Commission's stance on financial neutrality and the findings from Newhouse and colleagues, it may be appropriate to have a discussion about whether CMS should continue using data from FFS beneficiaries to calibrate the CMS–HCC model or switch to using data from MA enrollees. An argument for continued use of FFS data is that it is consistent with a policy of financial neutrality, and financial neutrality encourages care to be provided in the sector where it can be provided more efficiently (MA or FFS). Under financial neutrality, when plans are able to provide care at a lower cost than FFS Medicare, they may be able to offer enhanced benefits that are more attractive to beneficiaries than FFS Medicare. Alternatively, if plans cannot provide care at a lower cost than FFS Medicare, they may not be able to offer benefits that are competitive with FFS Medicare. An argument for use of MA data is that costs incurred by MA plans may differ from costs for FFS Medicare—perhaps because of different risk profiles between sectors or because of different models of care. To the extent these cost deviations occur between sectors, MA payments that reflect the cost of efficient providers require risk adjustment calibrated on data from MA enrollees.

Differences in coding between FFS Medicare and the MA program

MA plans have an incentive for providers to code their enrollees' conditions as completely as possible because MA risk scores and payments strongly depend on each enrollee's conditions. The incentive to code conditions is present but not as strong in FFS Medicare because FFS payments often depend on the services provided rather than beneficiaries' conditions.

This discrepancy in incentives between the MA program and FFS Medicare may be reflected in analyses by CMS and the Government Accountability Office (GAO), which found differences in diagnostic coding intensity between the two sectors (Centers for Medicare & Medicaid Services 2009, Government Accountability Office 2012). In response to its finding, CMS has reduced risk scores of MA enrollees by 3.4 percent in 2010, 2011, and 2012. However, GAO asserts that CMS has underestimated the magnitude of the greater coding intensity in the MA program by at least 1.4 percentage points and by as much as 3.7 percentage points. Statutory adjustments in the Patient Protection and Affordable Care Act of 2010 (PPACA) are consistent with the findings of the GAO. Starting in 2014, PPACA requires CMS to reduce MA enrollees' risk scores by an amount greater than 3.4 percent in each year, unless CMS begins using diagnosis and cost data from MA enrollees to estimate the CMS–HCC model. ■

Endnotes

1 CMS arrived at which HCCs to retain and how many to retain by balancing several competing considerations, including data collection burden, predictive power, whether to retain rare high-cost conditions, and retaining only well-defined, clinically coherent conditions (Pope et al. 2004).

2 It is not clear when CMS intends to begin using the data from MA enrollees to estimate the CMS–HCC model.

3 When we estimated the model using regression analysis, we used zero conditions as the basis of comparison, so we excluded that variable from the regression.

4 Financial neutrality can be obtained only if coding of conditions is the same in FFS Medicare and the MA program. Research by CMS and the Government Accountability Office (GAO) indicates that differences in coding exist between these two sectors (Centers for Medicare & Medicaid Services 2009, Government Accountability Office 2012). CMS has made adjustments to account for these differences, but GAO believes the adjustments are too small.

References

Brown, J., M. Duggan, I. Kuziemko, et al. 2011. *How does risk selection respond to risk adjustment? Evidence from the Medicare Advantage program.* NBER working paper 16977. Cambridge, MA: National Bureau of Economic Research.

Centers for Medicare & Medicaid Services, Department of Health and Human Services. 2009. Announcement of calendar year (CY) 2010 Medicare Advantage capitation rates and Medicare Advantage and Part D payment policies. Note from Jonathan D. Blum, acting director, Center for Drug and Health Plan Choice, and Paul Spitalnic, director, Parts C & D Actuarial Group, Office of the Actuary, to Medicare Advantage organizations, prescription drug plan sponsors, and other interested parties. April 6.

Frogner, B. K., G. F. Anderson, R. A. Cohen, et al. 2011. Incorporating new research into Medicare risk adjustment. *Medical Care* 49, no. 3 (March): 295–300.

Government Accountability Office. 2012. *Medicare Advantage: CMS should improve the accuracy of risk score adjustments for diagnostic coding practices.* Government Accountability Office report GAO–12–51. Washington, DC: GAO.

Medicare Payment Advisory Commission. 1998. *Report to the Congress: Medicare payment policy, volume II.* Washington, DC: MedPAC.

Medicare Payment Advisory Commission. 2000. *Report to the Congress: Improving risk adjustment in Medicare.* Washington, DC: MedPAC.

Medicare Payment Advisory Commission. 2001. *Report to the Congress: Medicare payment policy.* Washington, DC: MedPAC.

Medicare Payment Advisory Commission. 2002. *Report to the Congress: Medicare payment policy.* Washington, DC: MedPAC.

Medicare Payment Advisory Commission. 2004. *Report to the Congress: Medicare payment policy.* Washington, DC: MedPAC.

Medicare Payment Advisory Commission. 2005. *Report to the Congress: Issues in a modernized Medicare program.* Washington, DC: MedPAC.

Newhouse, J. P., M. B. Buntin, and J. D. Chapman. 1997. Risk adjustment and Medicare: Taking a closer look. *Health Affairs* 16, no. 5 (September–October): 26–43.

Newhouse, J. P., J. Huang, R. J. Brand, et al. 2011. The structure of risk adjustment for private plans in Medicare. *American Journal of Managed Care* 17, no. 6 (June): e231–240.

Pope, G. C., J. Kautter, R. P. Ellis, et al. 2004. Risk adjustment of Medicare capitation payments using the CMS–HCC model. *Health Care Financing Review* 25, no. 4 (Summer): 119–141.

Pope, G. C., J. Kautter, M. J. Ingber, et al. 2011. *Evaluation of the CMS-HCC risk adjustment model.* Report prepared by RTI International for the Centers for Medicare & Medicaid Services. Baltimore, MD: CMS.

Song, Y., J. Skinner, J. Bynum, et al. 2010. Regional variations in diagnostic practices. *New England Journal of Medicine* 363, no. 1 (July 1): 45–53.